Date: May 31, 2001

From: Mathematical Statistician (Gerry Gray) HFZ-542
Division of Biostatistics, OSB

Subject: Statistical Review for PMA P010015, Medtronic InSync Cardiac Resynchronization System for congestive heart failure.

To: Mitchell Shein (HFZ-450)
Division of Cardiovascular and Respiratory Devices, ODE
Through: Director, Division of Biostatistics, OSB ____

INTRODUCTION

The Medtronic InSync atrial synchronous biventricular (BV) pacing device is for the treatment of chronic heart failure (CHF) patients. This treatment of CHF via pacing is called cardiac resynchronization (CR) therapy.

The investigational portions of the InSync system consist of the model 8040 pulse generation device, the model 9980E software, the model 9790 programmer, and the Attain (model 2187 unipolar and model 2188 bipolar) internal left ventricular (LV) leads. The system also uses previously approved right ventricular and right atrial leads, and may use a model 4965 epicardial LV lead.

The original design was a three-month parallel trial (called MIRACLE, for "Multicenter InSync Randomized Clinical Evaluation"), in which patients were randomized to either "CR therapy on" to "CR therapy off". The original IDE was approved in 1998 and the sponsor went forward with the trial. After this trial was underway, FDA moved to require 6 month mortality data from all CHF trials; all subsequent submissions had this requirement. Thus the sponsor re-designed the MIRACLE trial as a parallel study with 6 month follow-up. Forty four centers participated in this trial (39 U.S., 5 Canada).

There were three co-primary endpoints identified in the IDE: NYHA classification, Minnesota Living with Heart Failure Quality of Life (QOL) questionnaire score, and 6-minute hall walk distance. The Hochberg procedure was used to adjust for potential p-value inflation due to multiplicity effects. Thus the MIRACLE trial would be a (statistical) success for effectiveness if any one of the three endpoints was statistically significant at alpha = 0.0167, if any two were significant at alpha = 0.025, or if all three were significant at alpha = 0.05.

The sponsor claimed in the IDE that the following differences in effectiveness were clinically significant:

- 13 points or more on the QOL score

- ½ class or more for NYHA class
- 50m or more for the hall walk.

The primary safety endpoints were InSync system survival and freedom from generator-related complications.

The study was also designed to demonstrate safety and effectiveness of the Attain model 2187/2188 leads. Primary endpoints for the leads include pacing threshold, implant success rate, and lead-related complications.

Secondary endpoints for BV pacing identified in the protocol include QRS width, peak VO_2, echocardiographic indices, plasma neurohormone levels, and patient survival.

SAMPLE SIZE AND PATIENT ACCOUNTABILITY

The sponsor separately calculated the required sample sizes to demonstrate the following (required sample sizes in parentheses):

- 95% lower confidence bound ≥ 90% for generator-related complications at six months (170 patients)
- 95% lower confidence bound ≥75% for lead-related complications at six months (170 patients)
- 95% lower confidence bound ≥ 80% for implant success (224 patients)
- 95% lower confidence bound ≥ 70% for system survival at 6 months (212 patients)
- 95% upper confidence bound ≤ 3.0v for lead voltage threshold at 6 months (83 patients)
- significant difference (alpha = 0.0167) between treatment and control in improvement in QOL score at 6 months (224 patients)
- significant difference (alpha = 0.0167) between treatment and control in improvement in NYHA class at 6 months (180 patients)
- significant difference (alpha = 0.0167) between treatment and control in improvement in 6-minute hall walk at 6 months (172 patients)
- significant difference (alpha = 0.05) in peak VO_2 (170 patients)
- no significant difference (alpha = 0.05) in mortality (using simple test for equivalence of proportions, with delta = 15%; 224 patients)

Based on these calculations, the sample size was set at 224 patients followed for six months.

Note that the sample size calculation for the three primary effectiveness endpoints uses the most conservative adjustment (i.e. alpha = 0.0167 using the Bonferroni procedure) for multiplicity. Thus this study may have been somewhat overpowered in the sense that it has high power to detect differences (reject null hypotheses) that may not be clinically significant.

Implants were attempted on a total of 567 patients, of which 527 were successful. 523 patients were randomized (262 control, 261 CR treatment). Of the 262 control patients, 38 were followed under the original 3-month protocol and 224 were followed under the revised 6-month protocol. Of the 261 CR patients, 33 were followed under the original 3-month protocol and 228 were followed under the 6-month protocol. Paired data (baseline and 6 months) for the primary effectiveness endpoints are available on 117 control and 124 CR patients.

BASIC CHF RESULTS

Both the CR and the control arm showed some improvement for all three of the primary effectiveness endpoints. The improvement in the CR arm was statistically significantly greater (using a Wilcoxon rank sum test) for all three of the endpoints. The device met the proposed clinical success criteria for NYHA classification, but did not meet the clinical success criteria for QOL score or 6-minute hall walk.

- For NYHA classification, the CR arm improved from a median of 3 at baseline to a median of 2 at 6 months, while the control arm median stayed constant at 3. This difference is statistically significant at $p < 0.001$. The 1 class difference in NYHA improvement is also greater than the specified clinically important difference in improvement.
- For the QOL score, the CR arm improved from a median of 60.0 at baseline to a median of 41.0 at six months, while the control arm improved from a median of 57.0 to a median of 47.0 over the same time period. This difference in improvement is statistically significant ($p = 0.017$). The 9 unit QOL improvement difference is less than the specified clinically important 13 units.
- For the 6-minute hall walk, the CR arm improved from a median 317 meters at baseline to a median 380m at 6 months, while the control arm improved from a median of 305m to a median of 321m over the same time period. This difference in improvement is statistically significant at $p = 0.004$. The 47m difference in 6-minute hall walk is slightly less than the specified clinically important 50 meters.

Thus the InSync device has met the IDE criterion for statistical success (with alpha as specified by the Hochberg procedure) for the primary effectiveness endpoints. However, it met only one of the three proposed clinical success criteria (NYHA class).

Using the data provided by the sponsor, individual patient results for the three primary endpoints are shown in Figures 1 – 3.

Figure 1. NYHA improvement (baseline to 6 months) for the two treatment groups.

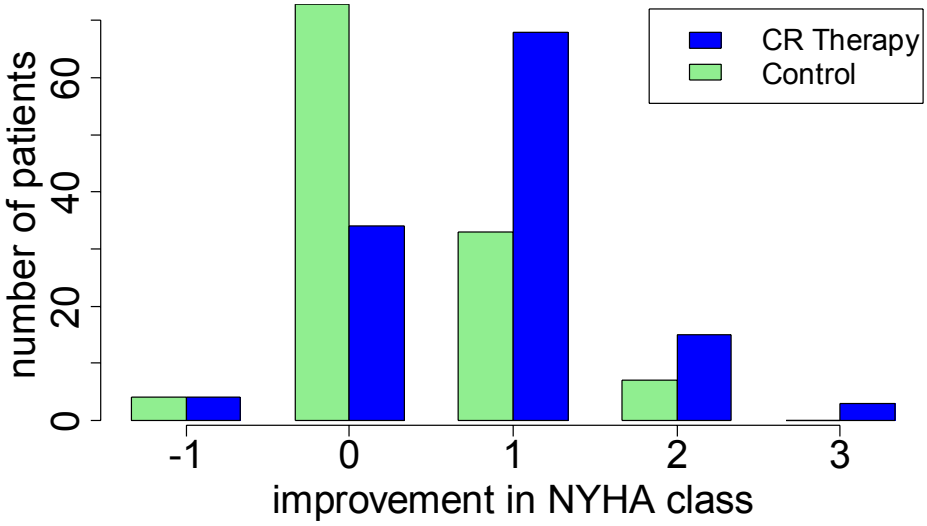


Figure 2. QOL scores for the two treatment groups. Line segments indicate individual patients, thick lines show the mean scores at each time point.

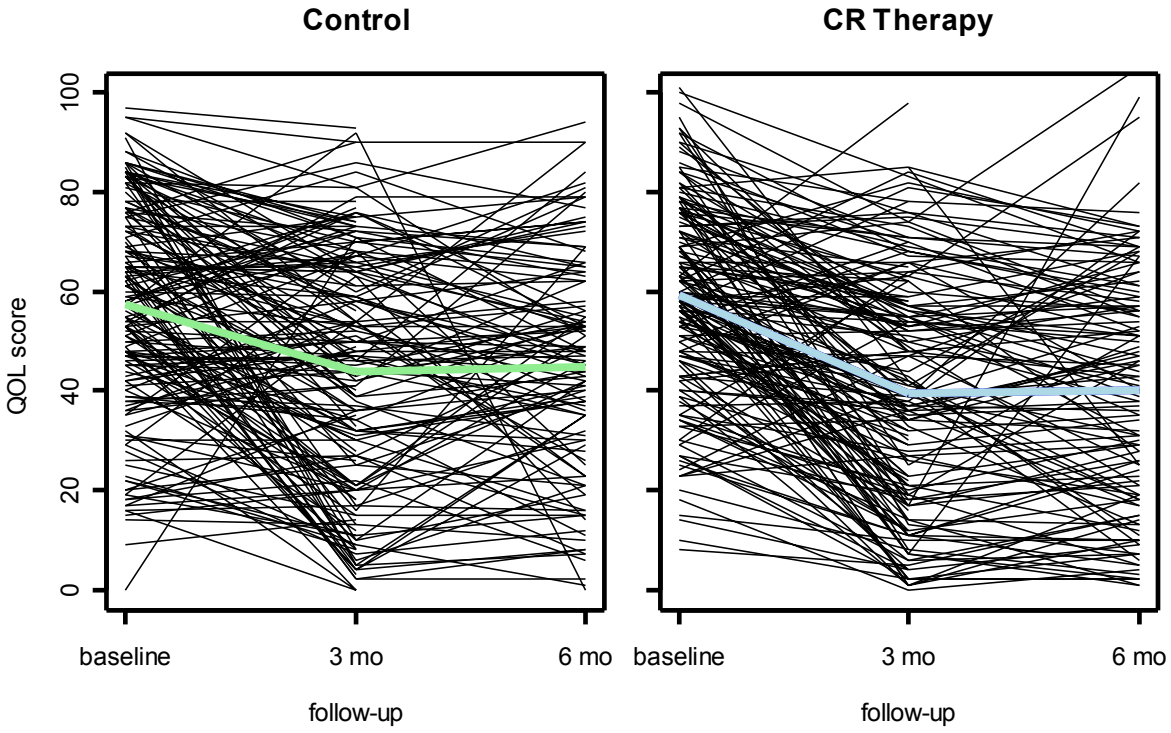
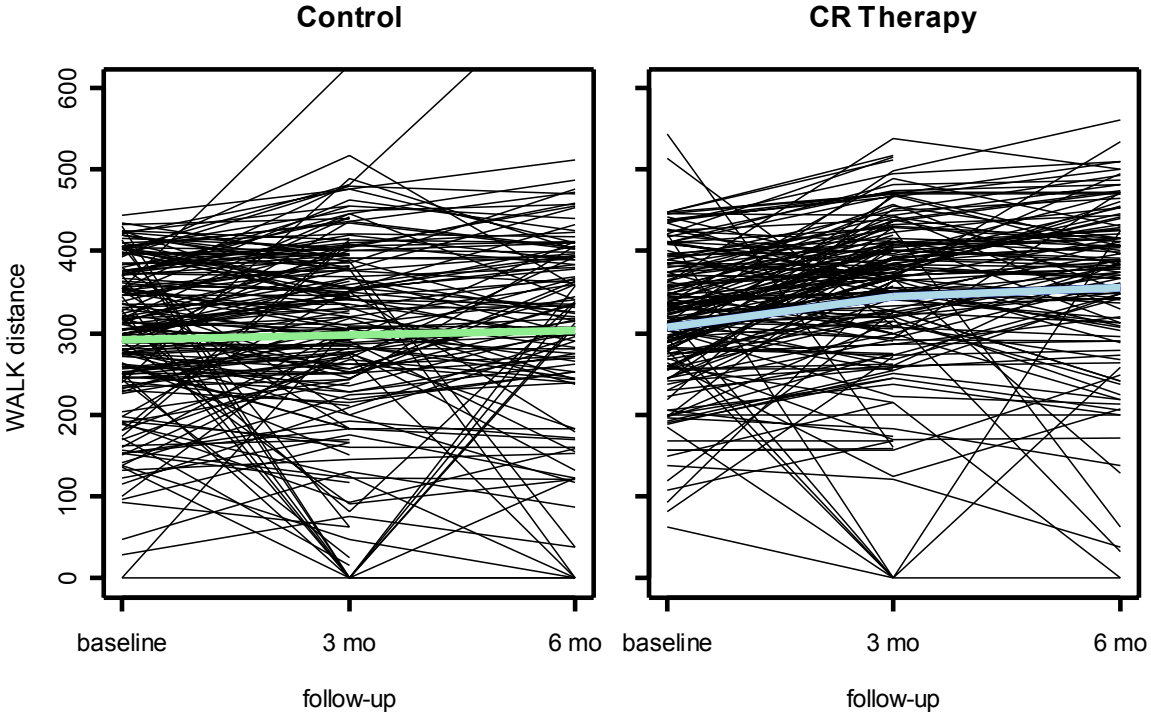


Figure 3. Hall walk distances for the two treatment groups. Line segments indicate individual patients, thick lines show the mean scores at each time point.



There were 11 deaths in the control arm and 12 in the CR arm during the first 6 months of follow-up (with patients from the original protocol censored at 3 months). Thus the mortality estimates (Kaplan-Meier) for the two groups (93.8% for the control arm and 94.3% for the CR arm) were virtually the same.

There were no generator-related complications during the first 6 months of treatment, thus the estimated complication-free rate is 100.0% (229/229; using the entire group of 527 successful implants), with an exact 95% confidence interval (c.i.) of [98.4%, 100.0%]. Thus the model 8040 generator meets the IDE criterion of a 90% or greater complication-free rate.

There were a total of 65 system-related complications in 48 of the 527 successfully implanted patients during the first 6 months of follow-up. The Kaplan-Meier estimate of freedom from system-related complications is 89.2%, with a 95% c.i. of [85.8%, 91.8%]. Thus the InSync system easily meets the IDE criterion of 70% or greater freedom from complications at 6 months.

BASIC LEAD RESULTS

The implant success rate for the Attain 2187 and 2188 leads was 92.9% (527/567), with an exact 95% c.i. of [90.5%, 94.9%]. Thus the lead implant success rate easily meets the IDE criterion of an 80% lower limit.

The Attain LV lead complication-free rate (Kaplan-Meier) was 92.4% at 6 months, with a 95% c.i. of [89.4%, 94.5%] (there were a total of 42 LV lead-related complications in 34 of 527 implanted patients). Thus the lead implant complication-free rate easily meets the IDE criterion of a 75% lower limit.

The Attain LV lead voltage threshold increased from a mean of 2.0v at baseline to 2.7v at one month, then gradually declined to 2.4v, 2.3v, and 1.9v over the 3, 6, and 12 month follow-up visits, respectively. The 95% confidence limits for voltage at 6 months is [2.1v, 2.4v]. Thus the Attain LV lead meets the IDE criterion of an upper 95% c.i. of 3.0v or less at 6 months.

There were 15 instances of complications described as "elevated thresholds" in Table 53 of the submission (5 resulted in lead replacement, 9 resulted in lead repositioning, and 1 resulted in invasive lead removal). Of these complications, 11 were in the CR group and 4 were in the control. Although this is not statistically significant, it should be commented on by the clinician. Specifically, I am wondering if there is any reason to believe that those patients with CR therapy "on" would have any increased risk of elevated thresholds.

If elevation of the voltage threshold above a certain point is classified as a complication, this raises two questions:

- Are these elevated thresholds included as part of the voltage threshold data?
- Should the performance of the LV leads be described by tolerance intervals (which characterize the proportion of leads with voltage thresholds above some value) instead of confidence intervals (which characterize the variability of the mean voltage threshold)?

Using the voltage threshold data provided by the sponsor, the 90% one-sided tolerance interval to contain 95% of the population is [0, 4.6v]. Thus we can say that we are 90% sure that 95% of all future leads will have a 6-month threshold below 4.6v.

ADDITIONAL ANALYSES

The sponsor carried out several additional analyses to check for gender differences and to model the primary outcomes of the study.

In the gender analysis, there were no differences between males and females for generator complications, LV lead complications, system complications, or LV lead pacing

thresholds. There was a significant difference in implant success rates (95% for males vs. 89% for females, p = 0.008). Further analyses indicated that the difference was explained by the increased QRS width for females (168.8 ms vs. 165.1 ms).

For the primary effectiveness endpoints, it appears that women show a greater improvement in QOL score than men. The other two effectiveness endpoints gave similar results for men and women.

To model the primary outcomes of the study, the sponsor used the baseline measurements of heart failure etiology, gender, age group, race, QRS width, ejection fraction, LV lead location, and bundle branch block type. Change in NYHA score was not associated with any of these variables in univariate models for each variable. Change in QOL score was significantly related to gender, LV lead location, and randomization group (CR therapy vs. control) in a multivariable model. Change in 6-minute hall walk was significantly associated with age group (age cut at 65 years) and randomization group in a similar model.

As a response to FDA requests, the sponsor carried out analyses of the primary effectiveness endpoints stratified by QRS width, LV ejection fraction, and baseline QOL, NYHA, and hall walk distance. These post-hoc subgroup analyses should be treated as exploratory (i.e. the p-values produced by these subgroup analyses are not valid).

- For QRS width the sponsor provided separate analyses for QRS width <=150ms, 151–170ms, 171–190ms, and >=190ms.
- For LV ejection fraction the sponsor provided separate analyses for EF <=15%, 16–25%, and >=25%.
- For NYHA class, the sponsor provided separate analyses for baseline class III and class IV patients.
- For hall walk, the sponsor provided separate analyses for baseline distances of <=224m, 225–299m, 300–374m, and 375–450m.
- For QOL score, the sponsor provided separate analyses for baseline scores of <=44, 45–59, 60–74, and >=75.

None of these subgroup analyses produced any striking results. In general, the direction and magnitude of the difference in improvement was similar across groups. Some reversals occurred when the sample sizes were small, but this is not surprising.

Using the data provided by the sponsor, I fit several logistic regression models for "success" (defined variously as meeting all 3 or any 2 of 3 clinical success criteria) using as predictors age, sex, QRS width, LV ejection fraction, baseline QOL, baseline hall walk, baseline NYHA, and treatment arm. Age, LV ejection fraction, and treatment arm were consistently significant predictors of success (older patients and those with lower LV ejection fractions had less chance of success).

BASELINE DIFFERENCES AND POOLING

The sponsor presents a summary of the baseline and demographic characteristics of the two arms in Table 22. There were no significant differences (in 14 variables) except for diastolic blood pressure ($p = 0.032$). The blood pressure difference of 1.9 mmHg was judged to be not clinically important.

Appendix V contains a summary of implant results by center. With the large number of centers in the trial, a formal statistical test for centers is not very useful. However, none of the centers had an obvious difference in implant success, and an exact test I carried out for difference between centers was not significant ($p = 0.10$).

SUMMARY

The InSync Cardiac Resynchronization System for congestive heart failure appears to have met all the statistical safety and effectiveness criteria set forth in the IDE. The device met one of the three specified clinical success criteria (NYHA class).

Since the sponsor provided an electronic copy of the data, several minor questions that came up during this review were easily answered without requiring further analyses by the sponsor.

Gerry Gray, Ph.D.

cc: Jim Dillard, (HFZ-450)
Donna-Bea Tillman, (HFZ-450)
Miriam Provost, (HFZ-450)
BIMO (HFZ-310)
Medical Device File
Board File